DataONE Webinar Series

Incentives, Challenges, Barriers: Exploring social, institutional and economic reasons for sharing data

Jens Klump
Heidi Laine
Fiona Murphy
Enable new science and knowledge creation through universal access to data about life on earth and the environment that sustains it

DataONE network of Member Nodes:
www.dataone.org/current-member-nodes

DataONE Search tool:
https://search.dataone.org
Lesson 1: Data Management

The world of data around us

The data deluge has created a surge of information that needs to be well-managed, discoverable, and accessible. The amount of available storage is not keeping pace with the amount of data being produced.

Why manage data:
- The researcher perspective
  - Keep yourself organized -> find your files!
  - Track your processes for reproducibility
  - Better version control of data
  - More efficient data quality control
  - More backups to avoid data loss
  - Format your data for reuse by yourself & others
  - Document your data for understandability & reuse
  - Prepare it to share it & gain credibility and recognition for your scientific efforts

Data management facilitates sharing and reuse.

Data Reuse Example
Researchers reused and aggregated data from several different sources to determine migration routes for specific bird species.

The Case for Data Management

If data are:
- Well-organized
- Documented
- Preserved
- Accessible
- Verified as to accuracy & validity

The results are:
- High quality data
- Data that is easy to share and reuse
- Citation & credibility to researcher
- Cost savings to further science

Lesson 2: Data Sharing

Addressing data sharing throughout the data lifecycle

Describe data content, character, and process. Deposit it in a location from which it can be accessed. Preserve it for formats & on media good for long term. Publish it so others can discover it. (Adria Richards)

Lesson 3: Data Management Planning

The Data Management Plan (DMP)

A DMP outlines what you will do with your data during and after you complete your research project. It is a formal but evolving document in which you lay out a plan to ensure that your data will be safe for the present and the future.

Component 1: Info about data & data formats and using the data, including spatial and temporal lay out a plan to ensure that your data will be safe for the present and the future.

Component 2: Metadata content & format
- What metadata are needed?
- How will metadata be created and/or captured?

Component 3: Policies for access, sharing, re-use

A tool to help you create your DMP: the DMPTool

Component 4: Long-term storage & management

The value of data sharing

The amount of available storage is not keeping pace with the amount of data being produced. The data deluge has created a surge of information that needs to be well-managed, discoverable, and accessible.

Information vs. Available Storage

Causes of data loss
- Natural disasters
- Facilities infrastructure failures
- Storage failure
- Server hardware or software failure
- Application software failure
- Human errors
- Malicious attacks
- Format obsolescence
- Loss of competencies
- Lost of funding
- Loss of institutional commitment

Costs of not doing data management can be very high!

The Data Lifecycle

The stages through which well-managed data passes from project inception to conclusion.

Local contact information

#DWS2016
www.DataONE.org
DataONE Webinar Series
www.dataone.org/webinars
Upcoming Webinar Event
www.dataone.org/upcoming-webinar
Previous Webinar Events (Recording and Discussion)
www.dataone.org/previous-webinars

#DWS2016
@DataONEorg
Upcoming Webinar Event
www.dataone.org/upcoming-webinar

How not to collect data: organizing data for long-term use and re-use

October 11
Stephanie E. Hampton
Washington State University
If you attending as part of a group, please enter the number of people listening within the “questions” box. Thanks!
Incentives, Challenges, Barriers: Exploring social, institutional and economic reasons for sharing data

Jens Klump  
CSIRO

Heidi Laine  
University of Helsinki

Fiona Murphy  
University of Reading
Data as Social Capital and the Gift Culture in the Scholarly Community

Jens Klump | OCE Science Leader Earth Science Informatics
13 September 2016
Why am I interested in data sharing?

- I am a geochemist, my field of research is Earth Science Informatics.
- Research data infrastructures have been part of my work since 1999.
- When I switched from marine geology to limnology I was puzzled by the difference in attitudes towards data sharing.
- Over the years I made some observations in this respect.

I am not a sociologist.
Why do communities behave so differently?

South Atlantic (Namibia)  |  Lake Baikal (Russia)
Are we getting the incentives right?
Waiting at the watering hole

- Sometimes waiting at the watering hole can be a successful strategy.
- The art is to identify suitable watering holes.
- Which resources do researchers need to access for their distinction gain?
- This is not only an opportunity to coerce compliant behaviour but also to develop better services for researchers.
Reputation Economy

Recognition —> Reception by Peers —> Publication —> Discussions —> Equipment —> Data —> Grant —> Recognition

Data as Social Capital  |  Jens Klump
Mineral Resources
Jens Klump
Science Leader Earth Science Informatics

t  +61 8 6436 8828
e  jens.klump@csiro.au
AFRAID OF SCOOPING?
Case Study on Perceived vs. Actualized Risks of Sharing Research Outputs

HEIDI LAINE / heidi.laine@helsinki.fi / @heidiklaine

Creative Commons Attribution 4.0 International License.
Misappropriation: unlawful presentation of another person’s result, idea, plan, observation or data as one’s own research

- Finnish Advisory Board on Research Integrity: Responsible conduct of research and procedures for handling allegations of misconduct in Finland - RCR guidelines, available at http://www.tenk.fi/en/responsible-conduct-research-guidelines
RESEARCH QUESTIONS

**Descriptive level:** How two pioneering open collaboration research projects coped with the risk of research misconduct?

**Interpretive level:** What allowed the researchers of the two case projects to ignore the risk of research misconduct?
Open Research Swarm / Social Media for Citizen Participation (SOMUS), 2007-2010

NMR Lipids Project (NMRLP), 2013-ongoing
### RADICAL OPENNESS

<table>
<thead>
<tr>
<th><strong>ORS / SOMUS</strong></th>
<th><strong>NMRLP</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Open collaborative writing of research proposal to a funding call</td>
<td>Participants required to give credit for ideas also from “hallway” discussions</td>
</tr>
<tr>
<td>Funding allocated to an open online community</td>
<td>Data shared via Github without embargo</td>
</tr>
<tr>
<td>Open collaborative writing of conference papers, meeting notes, wiki</td>
<td>Co-authorship determined by input through blog comments</td>
</tr>
</tbody>
</table>
• Research ethics and integrity training for researchers is vital also for advancement of open science.
• Multidisciplinarity fosters innovation: get physicists reading philosophy, engineering scientists working with sociologists.
• Data citation practices and data publication workflows should be implemented to ensure that openness pays of.
• Lack of gender balance indicative of women researcher’s lower tolerance of risk? Targeted efforts for encouraging women to share.
• Pioneers should be recognised and rewarded with funding and positions!
• Openness as an empowering experience should be promoted.

LESSONS TO BE LEARNED?
Longer version of this presentation: bit.ly/fearofscooping

Paper will be submitted for the Data Science Journal SciDataCon special collection and will be available as a pre-print on zenodo.org as of 1 October 2016.

Follow or contact me:
@heidiklaine / www.thehonestbrokerblog.org / heidi.laine@helsinki.fi
Giving Researchers Credit for their Data (or ‘Data2Paper’) 

Fiona Murphy, Neil Jefferies, Anusha Ranganathan, Thomas Ingraham, Hollydawn Murray 

DataOne Webinar, Denver, 13 September 2016
The Business Case

@dupuisj @Science_Open Hey, probably can't make it tonight. Fighting journal submission systems. Have a great one though :)

@andre PAPER Hey, need to submit a systems biology paper with lots of modelling data but the journal submission system will only accept Excel or PDF files.

@pat S chir is data should not be made available upon acceptance to a journal, they should be made available upon submission.

Susan Johnston
Submission to journal took several hours with limitations on size of supplementary data. Re-uploading to @biocorexpreprint took about 5 mins.

Anne Carpenter
Our supplemental data just crashed the journal's submission servers. Whoopsies. #BigData

Benjamin Saunders
It's not the most frustrating part of science, but navigating the online submission system of a journal has to break the top 10 for me.

Katherine James
Well, this journal submission form is utterly delightful, I'm not losing the will to live at all. Honest... *curls up under desk and weeps*

Matthew Cobb
@alisonatkin Peer-reviewed. Journal submission systems arrggghh

@alisonatkin Peer-reviewed. Journal submission systems arrggghh

Ahh, come here journal submission site -- let me feel your warm embrace. What a lovely long form!
The App

- Researcher logs in to app via institutional repository
- Automated data paper compilation within the research space (*no journal submission systems*)
- Article packages submitted to publisher system via SWORD
- If published, ORCID updates the record

Title: Contextual and provenance metadata in the Oxford University Research Archive

Abstract: Context and provenance are essential for understanding the meaning and significance of an artefact. In this paper we describe how scholarly outputs deposited in a long-term data repository, the Oxford University Research Archive (ORA), are described with contextual information and provenance. In addition, the digital objects in ORA that act as proxies to the scholarly outputs are also described with contextual information and provenance. The ORA data model is presented, together with a description of the relationships in context.

About The Authors
- Tanya Gray Jones
  Institution: University of Oxford
  Faculty: The Bodleian Libraries

- Lucie Burgess
  Institution: University of Oxford
  Faculty: The Bodleian Libraries

- Neil Jeffries
  Institution: University of Oxford
  Faculty: The Bodleian Libraries

- Anusha Ranganathan
  Faculty: The Bodleian Libraries

- Sally Rumsey
  Institution: University of Oxford
  Faculty: The Bodleian Libraries

Bibliographic Details
- Publisher: University of Oxford
- Date published: 2015
- Copyright Date: 2015
- DOI: 10.5072/bodleian:nZaoXzEfm5

Identifiers
- Uris: uris:6083416-c8cc-4c89-a929-b7f83e980cc2
- DOI: 10.5072/bodleian:nZaoXzEfm5
Next Steps

- Feedback, UI/UX work
- Researchers – domain and social factors
- Outreach and trials
- Sustainable long-term strategy
- All suggestions welcome!
Thank you!

fionalm27@gmail.com
@DrFionaLM

Jisc: Research Data Spring Grant
DataONE Webinar Series

www.dataone.org/webinars

Upcoming Webinar Event

www.dataone.org/upcoming-webinar

Previous Webinar Events (Recording and Discussion)

www.dataone.org/previous-webinars

#DWS2016

@DataONEorg